

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

E7.4-10587

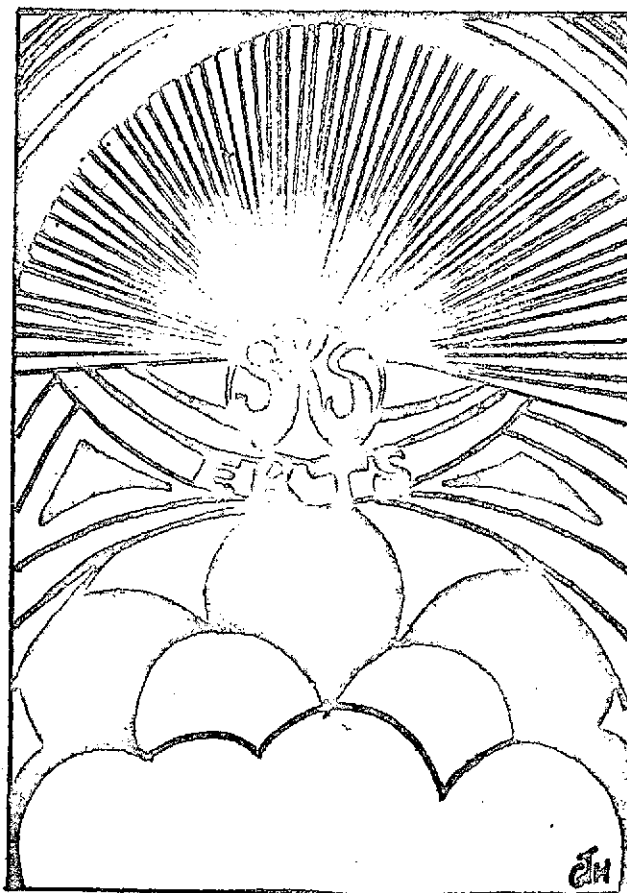
CR-138807

Type II Progress Report

January 8, 1974 - June 19, 1974

Crop Identification & Acreage

Measurement Utilizing ERTS Imagery 013



Principle Investigator

Donald H. Von Steen AG 328

(E7.4-10587) CROP IDENTIFICATION AND
ACREAGE MEASUREMENT UTILIZING ERTS IMAGERY
Progress Report, 8 Jan. - 19 Jun. 1974
(Department of Agriculture) 9 p HC
\$4.00

CSCL 02C G3/13

Unclass
00587

N74-28830

Table of Contents

	Page
Analysis of Aerial Photography	1
Ronald Steele	
Analysis of ERTS Data	2
William Wigton	
Software	5
Paul Hopkins and Fred Warren	
Microdensitometer	7
Ronald Steele	

Analysis of Aerial Photography

Scanning of the RC-8 high altitude photography from mission 208, flown by NASA on August 18, 1972, has been completed. This photography covers flightlines 3 and 10 in Kansas. Scanning is nearing completion from mission 211 which covers flightlines 3 and 5 in South Dakota. It was found that locating fields within segments and recording their coordinates were time-consuming tasks. Sketches of each segment were drawn showing each field within the segment. Corner coordinates of rectangular fields were then recorded on the sketch from the DCRS (digital coordinate readout system) on the microdensitometer system. For irregular shaped fields (non-rectangular), as many as 10 boundary coordinates were recorded. The segments were then scanned by the microdensitometer with an effective aperture of 240 microns square and the optical densities and percent transmission values were recorded on magnetic tapes for each of the four color filters (red, green, blue, and clear). The microdensitometer scanning data from the Kansas test sites (segments) has successfully been converted to a SAS (Statistical Analysis System) compatible format using the PDSCMS computer program. A computer program has been developed which assigns tract, field, and crop identifiers to each data observation (aggregate data from one pixel) on the basis of ground truth information. The program operates in conjunction with SAS to extract rectangular fields parallel to the microdensitometer scanning axes. Irregular shaped fields are subdivided into several rectangular fields parallel to the scanning axes. Field extraction has been completed for the Kansas segments, and the data is now in a format suitable for discriminant analysis.

Cost Analysis

The following is a breakdown of approximate time and cost involved in scanning the aircraft photography in Kansas, and converting the data into a format suitable for crop classification.

Activity

Average Time/Segment

Sketch segment and record field boundaries	37 min.
Microdensitometer Scanning	33 min.
Recording and Key punching input data for field extraction	40 min.
Total man hours	= 1.83 hours
Approx. cost/man hour	= \$4.50
Average cost/segment	= 1.83 x 4.50 = \$8.23

ADP costs on a per segment basis are as follows:

PDSCMS data conversion	\$12.00
Field extraction	\$ 9.00
Total	\$21.00

Thus, the average cost per segment for scanning the aircraft photography and converting the data into a format suitable for crop classification is \$29.23.

Analysis of ERTS Data

Idaho

Classification results in Idaho were poor. One reason for the poor results seemed to be a banding problem. Upon our request, NASA reprocessed Idaho frame 1035-17525 to remove the banding. Classification was done using the reprocessed tapes and nearly identical results were obtained as were reported previously.

Kansas

ERTS imagery for the Kansas study area was too cloudy to be useful prior to September 21, 1972. The study was made on September 21 and 22 imagery. The imagery for the area of interest in Kansas was divided by two consecutive ERTS passes, thus the training data was also divided. Twenty-two segments were in the September 21, imagery. Seven of these segments were hidden by clouds. Therefore, 15 segments were used as training and test data. The September 22 imagery contained 23 segments, one of which fell in a non-agricultural area. Two additional segments were selected containing sugar beets for training purposes.

The standard pointwise quadratic discriminant functions found in LARSYS were used for classification, with the added feature of unequal prior probabilities. Using four groups for classification with unequal prior probabilities, the weights used in Kansas were:

- | | | |
|-----|---------------|-----|
| 1). | Alfalfa | .03 |
| 2). | Pasture | .72 |
| 3). | Corn | .09 |
| 4). | Grain Sorghum | .16 |

These prior probabilities were computed from data gathered by the Statistical Reporting Service in early June 1972, for Crop Reporting District 7 in Kansas (area corresponding to the test site).

Since the time of year was not optimum, a visual inspection of the gray scale printout of MSS band 5 and ground truth was used to select training fields. Segments contained in the imagery for both portions of the test site were classified twice. For the first classification, fields which were partially harvested and those with a confusion of symbols on the gray scale printout were discarded. All identifiable fields were used for the second classification. The overall percentage of pixels correctly classified is shown below.

Date of Imagery	Number of Segments	Overall Classification Performance	
		Run 1*	Run 2**
September 21	15	91.2%	90.2%
September 22	24	75.5%	65.8%

* partially harvested fields, etc. discarded.

** all identifiable fields used.

Prior probabilities used corresponded to the entire Kansas test area. The ERTS imagery divided the test area. Better classification results would have been possible if the prior probabilities corresponded to the portion of the test site within each ERTS frame. However, development of proper weights for areas divided by consecutive ERTS passes presents additional problems.

The decrease in overall classification performance between the September 21 and September 22 imagery could be attributed to several factors.

- 1). The number of crops being classified was increased from four to seven. Increasing the number of crops will reduce the performance.
- 2). There was a confusion between most crops and pasture. This could have resulted from using late September imagery and the weight given to pasture was .72.

The correlation between acres and pixels were calculated. Coordinates of ground truth segments were carefully defined. The training data from each scene were used to classify the segments in that scene. The classified pixels in the two scenes were then combined (i.e. Tables 2 and 4 were combined) and correlations with known ground truth acreage were computed.

Correlations between acreage and pixels were calculated as follows:

Total acreage vs Total Pixel	$r^2 = .88$	$r = .94$
Pasture acreage vs Pasture Pixel	$r^2 = .84$	$r = .92$
Corn acreage vs Corn Pixel	$r^2 = .62$	$r = .79$
Grain Sorghum vs Grain Sorghum Pixel	$r^2 = .58$	$r = .76$

When pixels and acreage are this highly correlated, remotely sensed data is beneficial.

Signature extension was studied by compiling training statistics on one ERTS frame and using them to classify points in the adjacent frame. The overall percentage of pixels correctly classified on the September 21 imagery using statistics generated from the September 22 imagery was 85.5%. When the procedure was reversed, 49.0% of the pixels were correctly classified. The exercise was conducted as an experiment, and the authors do not imply signature extension could be applied in general. It has not been shown that in general one would expect good or bad results from such a practice. However, the results obtained here do show hope in the area of using training data in one frame to classify in others. The ability to use training data in more than one frame would be of great benefit when working with a very large area. Calibration changes between ERTS frames do create problems that must be dealt with.

County estimates were studied by first obtaining the border of Stevens County, Kansas on a gray scale map using MSS band 5. The area was then defined on punch cards and classified. Training data for classification was obtained from segments in the Crop Reporting District which contained Stevens County. Three of these segments were actually in Stevens County. A total of 410,505 pixels were classified which expands (1 pixel = 1.13 acres) to 463,871 acres compared to 466,565 acres reported for Stevens County in the 1964 Census of Agriculture. The prior probabilities which were applied were the same as those used earlier for Kansas. There is an indication of confusion between pasture and grain sorghum. Ways to use this data to produce a final estimate are being investigated.

South Dakota

In South Dakota, late September imagery was the only available imagery for our district because of clouds in earlier imagery. Classification results were poor. Examination of the Coincident Spectral Plot showed very little information in the ERTS data for the separation of the classes of interest. This late in the season, crops were classified as either oats or pasture.

The use of fields selected from gray scale printouts and ground truth did not improve classification and actually reduced the overall performance.

Software

PDSCMS

The program to convert microdensitometer data into SAS compatible observation has been modified to permit the use of a STACK option of output data sets. Basically, this permits the user to create up to 249 output data sets during a single run. Formerly, a separate job or job step was required for each data set being produced.

This option permits greater thru-put of microdensitometer data by requiring less JCL in the job setup, or many more data sets prepared in one day.

The PDSCMS program is described in the TYPE II progress report dated June 20 - December 19, 1973.

SIPFXS

SAS IMAGE PROCESSING: FIELD EXTRACTION by SEQUENCING is a new software program being developed in-house. It will take a SAS compatible data set produced by PDSCMS and assign serial numbers to data points corresponding to their 'field' position. The SAS system will then sort these points in ascending order, so that all the data for a given field will be physically contiguous in the data set.

The program is being designed to permit the processing of non-convex polygons with n-sides. Data points will be tagged with a sequence number, group and ident names for classification, and JES tract and field idents for ease in later processing using the SAS system.

SAS

The modifications for the SAS discriminant analysis system have been written by Barr and Goodnight at North Carolina State University. The modifications will permit the handling of relatively large data sets as well as several convenience options. These include saving and re-use of the mean vectors, separation of calibration and unknown data into separate data sets, prior probabilities, thresholding, and optional error listing of misclassified points.

We will be getting a pre-release version soon, and it should be generally available with the next release of the SAS system.

Penn State Classifier: Version III

We are in the process of installing the new version now. The first group of programs is nearly complete. The programs running are NMAP, ACLASS, TPINFO, and STATS. The SUBSET program was accidentally left off the tape copy sent us. We have asked for and received a new copy with SUBSET included.

The new tape that we received from Dr. Borden has a Quadratic classifier with prior probabilities. We will put this classifier up right away. Hopefully, this will give us in-house capability comparable with the LARSYS classifier.

RADFWMAP

A computer program (RADFWMAP) has been developed to produce grey scale "maps" from ERTS computer-compatible MSS bulk data tapes. Particular abilities of this program are:

- 1). The program will run from 1 1/2 to 4 times as fast as the Penn State NMAP program.
- 2). The user has the option of mapping from response bands 4, 5, 6, or 7.
- 3). The user can assign sets of up to 15 printable characters to be used in the mapping.
- 4). The program will compute a frequency tabulation of response values in any designated area of the tape, and use this tabulation to assign the printable characters to the classes of response values. Subject to constraints imposed by the frequency distribution, the proportion of data points in each class will be approximately equal.
- 5). Numbering of elements in the scan line is compatible with the Penn State and LARSYS system.

Microdensitometer

Scanning has been intermittently delayed due to stage runaway problems and testing procedures. The Perkin-Elmer Corporation has decided to replace the DCRS (Digital coordinate readout system) unit of the microdensitometer system in hopes of curing the stage runaway problem. This will result in an additional 10-15 day delay for installation and running acceptance tests.